

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»  
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ  
Кафедра інформаційної безпеки**

**«До захисту допущено»**

В.о. завідувача кафедри

\_\_\_\_\_ М.В.Грайворонський

(підпис)

“ \_\_\_\_\_ ” \_\_\_\_\_ 2019 р.

**Дипломна робота  
на здобуття ступеня бакалавра**

з напрямку підготовки 6.170101 «Безпека інформаційних і комунікаційних систем»  
на тему: “Застосування методів реідентифікації осіб для аналізу матеріалів систем  
відеоспостереження”

Виконала: студентка 4-го курсу, групи ФБ-51

(шифр групи)

Мерзлікіна Поліна Миколаївна

(прізвище, ім'я, по батькові)

(підпис)

Керівник к.т.н., доцент Родіонов А. М.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Консультант

\_\_\_\_\_

(назва розділу)

(посада, вчене звання, науковий ступінь, прізвище, ініціали)

(підпис)

Рецензент

\_\_\_\_\_

(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали)

(підпис)

Засвідчую, що у цій дипломній роботі немає  
запозичень з праць інших авторів без відповідних  
посилань.

Студент \_\_\_\_\_

(підпис)

Київ - 2019 року

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**  
**імені ІГОРЯ СІКОРСЬКОГО»**  
**ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**  
 Кафедра інформаційної безпеки

Рівень вищої освіти – перший (бакалаврський)

Напрямок підготовки 6.170101 «Безпека інформаційних і комунікаційних систем»

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри

\_\_\_\_\_ М.В.Грайворонський  
 (підпис)

« \_\_\_\_ » \_\_\_\_\_ 2019 р.

**ЗАВДАННЯ**  
**на дипломну роботу студенту**

Мерзлікіній Поліні Миколаївні  
 (прізвище, ім'я, по батькові)

1. Тема роботи

“Застосування методів реідентифікації осіб для аналізу матеріалів систем відеоспостереження”,  
 науковий керівник роботи Родіонов Андрій Миколайович, к.т.н., доцент.  
 (прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від « \_\_\_\_ » 2019 р. № \_\_\_\_\_

2. Термін подання студентом роботи \_\_\_\_\_

3. Вихідні дані до роботи

1. Архітектури та методи навчання глибинних нейронних мереж.
2. Вибірка на основі відеоматеріалів камери відеоспостереження
3. Мова програмування Python та бібліотеки для роботи з зображеннями.

4. Зміст роботи

1. Дослідити існуючі підходи застосування реідентифікації
2. Обрати методи виявлення та відслідковування, які можуть бути застосовані для роботи з відео
3. Проаналізувати специфіку вихідного відеоматеріалу
4. Виконати програмну реалізацію
5. Проаналізувати отримані результати

## 5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)

## 1. Презентація MS PP

6. Дата видачі завдання \_\_\_\_\_

## Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів дипломної роботи	Примітка
1	Отримання завдання		
2	Огляд стану предметної області		
3	Дослідження теоретичних основ методів реідентифікації		
4	Аналіз існуючих методів		
5	Аналіз вихідних даних		
6	Опанування обраних програмних засобів		
7	Виконання програмної реалізації		
8	Аналіз отриманих результатів		
9	Оформлення текстової та графічної документації		

Студент

\_\_\_\_\_  
(підпис)

Науковий керівник роботи

\_\_\_\_\_  
(підпис)

Мерзлікіна П. М.

(ініціали, прізвище)

Родіонов А. М.

(ініціали, прізвище)

## РЕФЕРАТ

Загальний обсяг роботи 49 сторінок, 17 рисунків, 3 таблиці, 20 бібліографічних найменувань.

Метою роботи є побудова методу реідентифікації осіб на відео з використанням нейронних мереж.

Об'єктом дослідження є відеоматеріали камер відеоспостереження.

Предметом дослідження є алгоритми виявлення, відслідковування та реідентифікації об'єктів на основі глибинного навчання.

Методами дослідження є моделі штучних нейронних глибинних мереж, програмні бібліотеки для роботи з зображеннями.

Результати роботи пропонують комбінацію методів для автоматизації обробки матеріалів систем відеоспостереження, що мають за основу застосування методів штучного інтелекту.

Виконання даної кваліфікаційної роботи дозволило отримати первісну статистику кількості осіб, оцінити якість застосування алгоритмів виявлення та відслідковування осіб для відеоматеріалу, та визначити осіб, які найчастіше з'являлись у полі зору відеокамери.

Ключові слова: реідентифікація осіб, виявлення об'єктів, відслідковування декількох об'єктів, глибинні нейронні мережі, машинне навчання

## **ABSTRACT**

Diploma work: 49 pages, 17 figures, 3 tables, 20 references.

The goal of the diploma thesis is to build a method for person re-identification on video by using neural networks methods.

The object of research is video materials from a surveillance video camera.

The subject of the research is human detection, multi-object tracking, and person re-identification algorithms, based on deep neural networks.

Research methods used in this thesis are deep neural network models and image manipulation programming libraries.

The results of the study offer a combination of methods based on artificial intelligence for automated analysis in the video surveillance systems.

The results of this qualification work allowed to receive primary statistics of a number of people in the video, make an evaluation of human detection and tracking algorithms performance on video material, and to identify most often found people in the camera field of view.

Keywords: person re-identification, object detection, multi-object tracking, deep neural networks, machine learning

## ЗМІСТ

Перелік умовних позначень, символів, скорочень і термінів.....	8
Вступ.....	9
1 Принципи побудови системи реідентифікації осіб.....	11
1.1 Загальні поняття.....	11
1.1.1 Приклади застосування.....	12
1.1.2 Обмеження при застосуванні.....	12
1.2 Опис типової системи Re-ID.....	13
1.2 Загальний алгоритм повторної ідентифікації особи.....	15
Висновки до розділу 1.....	17
2 Огляд підходів виявлення і відстеження людей на відео.....	19
2.1 Виявлення об'єктів.....	19
2.1.1 Методи виявлення об'єктів.....	19
2.1.2 Метрика оцінки виявлення об'єктів.....	22
2.2 Відстеження декількох об'єктів .....	24
2.2.1 Формалізація проблеми відстеження декількох об'єктів.....	26
2.2.2 Огляд класичних методів відстеження об'єкта .....	26
Висновки до розділу 2.....	27
3 Вибір архітектур необхідних нейронних мереж.....	29
3.1 Глибинні нейронні мережі.....	29
3.2 Вибір архітектури штучної нейронної мережі для виявлення людей .....	31
3.3 Вибір архітектури штучної нейронної мережі для відстеження людей .....	32
3.4 Застосування нейронних мереж для реідентифікації.....	32
Висновки до розділу 3.....	34
4 Розроблення програмної реалізації.....	35
4.1 Опис вхідних даних.....	35
4.2 Застосування готових програмних рішень.....	36
4.3 Власна програмна реалізація.....	37
4.3.1 Створення анотацій.....	37
4.3.2 Виявлення людей.....	38
4.3.3 Відслідковування людей.....	41

	7
4.3.4 Застосування реідентифікації.....	42
4.4 Аналіз отриманих результатів.....	43
Висновки до розділу 4.....	45
Висновки.....	46
Перелік джерел посилань.....	47

## **ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ І ТЕРМІНІВ**

AP — Average Precision

CNN — Convolution Neural Network, згорткова нейронна мережа

DBT — Detection Based Tracking

GPU — graphics processing unit, графічний процесор

HOG — Histogram of Oriented Gradients

MAP — maximal a posteriori

MOT — Multi-Object Tracking

OpenVINO — Open Visual Inference and Neural network Optimization toolkit

Re-ID — Re-Identification, реідентифікація

SORT — Simple Online and Realtime Tracking

SSS — Small Sample Size

YOLOv3 — You Only Look Once version 3



## ВСТУП

**Актуальність роботи.** У сучасному світі активно поширюються системи відеоспостереження, проте вони збирають настільки багато даних, що оператори перестають бути спроможними виконувати їх обробку. З розвитком комп'ютерного зору з'являються нові підходи для полегшення аналізу отриманої з камер інформації, дана робота направлена на їх застосування для практичних задач. Зазвичай якість відеоматеріалу з відеокамер є не високою через багато факторів, такі як вартість камер, розміщення на висоті, обмеження розміру відеозаписів для подальшого зберігання, тому такі підходи як розпізнавання людей по обличчю не є застосовними і необхідне застосування інших методів, реідентифікація осіб є одним із них.

**Мета дослідження.** Побудова методу реідентифікації осіб на відео з використанням нейронних мереж.

### **Завдання роботи.**

1. Дослідити існуючі підходи застосування реідентифікації.
2. Обрати методи виявлення та відслідковування, які можуть бути застосовані для роботи з відео
3. Проаналізувати специфіку вихідного відеоматеріалу
4. Виконати програмну реалізацію
5. Проаналізувати отримані результати

**Об'єкт дослідження.** Відеоматеріали камери відеоспостереження

**Предмет дослідження.** Алгоритми виявлення, відслідковування та реідентифікації об'єктів на основі глибинного навчання.

**Методи дослідження.** Моделі штучних нейронних глибинних мереж, програмні бібліотеки для роботи з зображеннями.

**Наукова новизна одержаних результатів.** Результати роботи пропонують комбінацію методів для автоматизації обробки матеріалів систем відеоспостереження, що мають за основу застосування методів штучного інтелекту.

**Практичне значення одержаних результатів.** Результати виконання даної кваліфікаційної роботи дозволили отримати первинну статистику кількості осіб та визначити осіб, які найчастіше з'являлись у полі зору відеокамери.

# 1 ПРИНЦИПИ ПОБУДОВИ СИСТЕМИ РЕІДЕНТИФІКАЦІЇ ОСІБ

## 1.1 Загальні поняття

Повторна ідентифікація особи (Person re-identification, Re-ID) — процес встановлення відповідності між зображеннями особи, зробленими з різних камер, або з однієї камери в різні моменти часу; призначення стабільного ідентифікатора різним екземплярам зображення особи [1].

Реідентифікація особи є фундаментальним завданням автоматизованого відеоспостереження і є предметом інтенсивного дослідження останні кілька років. За цей час було запропоновано велику кількість методів повторної ідентифікації, які ґрунтуються на розробці спеціальних характеристик зображень осіб, які здатні точно характеризувати кожного суб'єкта, та на створенні відповідних показників, за якими можна порівнювати знайдені характеристики кожної людини. Крім того, було запропоновано низку публічних наборів даних для створення однакових умов розвитку для наукового співтовариства [2].

Як задача повторна ідентифікація для людини є інтуїтивно зрозумілою, ми робимо це весь час без особливих зусиль. Наші очі і мозок навчені виявляти, локалізувати, ідентифікувати і пізніше повторно визначати об'єкти і людей у реальному світі. Повторна ідентифікація передбачає, що особа, яка раніше була виявлена, ідентифікується в наступній появі, використовуючи той самий унікальний дескриптор особи. Для зображень людей такий дескриптор ґрунтується на обличчі людини, її висоті та статури, одязі, кольорі волосся, зачісці, ході тощо. Оскільки процес розпізнавання в основному базується на зовнішньому вигляді суб'єктів, то припускається, що одяг залишається незмінним, і не передбачає розпізнавання облич або інших біометричних сигналів.

### **1.1.1 Приклади застосування**

- Відстеження на декількох камерах

Для відстеження особи з використанням декілька камер, ідентифікатор особи з другої камери необхідно отримати на основі інформації, отриманої від першої камери.

- Відстеження траєкторії

Якщо місця розташування камер відомі, виходячи з системи повторної ідентифікації, можна відстежити переміщення шляху людини від однієї точки до іншої.

- Засоби спостереження та безпеки

Наприклад, для відстеження підозрюваного на місці злочину

### **1.1.2 Обмеження при застосуванні**

Повторна ідентифікація людини є складною проблемою через візуальну неоднозначність і просторово-часову невизначеність зовнішнього вигляду людини в різних камерах. Ці труднощі часто доповнюються низькою роздільною здатністю зображень чи відеопотоками низької якості з великою кількістю незв'язаної інформації, що робить Re-ID проблемою, яка не має ефективного розв'язку в реальних середовищах і застосунках.

Методи реідентифікації мають два фундаментальні обмеження при застосуванні до реальних задач [3]:

#### **1. Малий розмір вибірки**

Тренувальні дані для навчання використовуваних нейронних мереж часто невеликі за обсягом (наприклад, декілька сотень осіб з декількома зображеннями) і набагато менші ніж типові розмірності векторів ознак. Відсутність достатньої кількості навчальних зразків відома як проблема малого розміру вибірки (Small Sample Size, SSS). Враховуючи те, що метричні методи навчання повторної ідентифікації спрямовані на мінімізацію дисперсії всередині класу, одночасно

максимізуючи міжкласові дисперсії, проблема SSS, таким чином, може зробити рішення неефективними.

## 2. Непристосовність

Існуючі методи повторної ідентифікації для створення моделі потребують досить великий обсяг навчальних даних, зібраних через процес ручного маркування, який вимагає багато часу. Ця схема першого маркування та натренування не є масштабованою для реальних застосунків, які вимагають розгортання на багатьох місцях спостереження з не анотованими даними. Збирання міток паралельно в процесі роботи застосунка є інкрементним, тобто додаткові мітки послідовно доступні для оновлення моделі з плином часу, тому модель повторної ідентифікації поступово зростає і безперервно адаптується до нових доступних актуальних даних. Існуючі способи повторної ідентифікації дозволяють лише повторне навчання моделі з нуля, що призводить до високих витрат на обчислювальну систему та затримки відповіді для користувача.

### 1.2 Опис типової системи Re-ID

Типова система Re-ID має два основні компоненти:

1. Знаходження унікального дескриптора особи (отримання моделі)
2. Порівняння двох виявлених моделей, на базі якого робиться висновок про їх відповідність або про невідповідність одна одній.

Типова система Re-ID в якості вхідних даних для виявлення об'єктів та створення дескрипторів може мати зображення або відео. Якщо на вхід подається зображення, особа має бути надійно виявлена і локалізована для точного отримання ознак.

Для того щоб визначити унікальний дескриптор для зображення людини, необхідна можливість автоматичного виявлення і відстеження людей у зображеннях або відео. Для автоматизації кожного компонента необхідно виконати ряд завдань, які мають свої власні виклики і впливають на загальну складність реалізації Re-ID.

Якщо доступно декілька зображень, для забезпечення впевненості у тому, що отримані ознаки на різних зображеннях належать певній людині, необхідною є можливість встановлення відповідності між виявленими суб'єктами через кадри. Цей процес називається відстеженням і забезпечує послідовну мітку для кожного суб'єкта в декількох кадрах. Таким чином, декілька екземплярів зображення конкретної особи можуть бути використані для вилучення ознак і подальшого генерування дескриптора для використання при повторній ідентифікації. Відстеження кількох об'єктів (Multiple Object Tracking, MOT) в межах однієї камери було широко досліджено, протягом останніх двох десятиліть запропоновано багато алгоритмів, проте стійке відстеження при різних умовах спостереження залишається відкритою проблемою.

Припускаючи, що точне виявлення особи та відстеження з використанням однієї камери можливе, першим кроком у Re-ID є знаходження візуального дескриптора людини. Важко забезпечити високу якість візуальних даних, оскільки такі чинники, як роздільна здатність, частота кадрів, умови та кут зйомки зображення широко варіюються і не завжди можуть контролюватися. Люди на результуючому зображенні можуть частково або повністю перекривати один одного через натовп чи сторонні об'єкти. Отримання надійного дескриптора залежить від наявності якісних спостережень. Неправильні виявлення та неправильна оцінка траєкторії вводять помилки в процес вилучення та генерації дескрипторів, які безпосередньо впливають на якість Re-ID.

Найпростішим і найбільш очевидним дескриптором особи, який легко можна отримати з відеоданих, є зовнішній вигляд, який характеризується такими ознаками як колір, текстура і форма. Однак ці дескриптори не є унікальними і схильні до змін. Дескриптори кольору/текстури недостатньо описові і різко змінюються внаслідок зміни освітлення, варіацій поз або кута огляду чи змін масштабу, притаманних системам з декількома камерами. Зчленований характер людського тіла призводить до різноманітності у формах силуетів, а різні геометрії камери роблять дескриптори форми менш дискримінаційними. Якщо два зображення знімаються лише через кілька хвилин або годин, то дескриптори на основі

зовнішнього вигляду можуть виявитися достатніми для використання в системі Re-ID. Ця ідея базується на припущенні про те, що люди, швидше за все, будуть в одному і тому ж одязі, оскільки одяг є визначальним фактором зовнішнього вигляду. Це не означає, що одяг є найкращим дескриптором у даному сценарії, проте він є прийнятним. Такий тип Re-ID сценарію називається короткоперіодним Re-ID, та довгогоперіодним для зображень/відео, які беруться за кілька днів чи місяців. Часова різниця між зображеннями є фактором складності Re-ID.

Порівняння дескрипторів людини є складним завданням через невизначеність, пов'язану з можливою відсутністю завчасно відомих просторово-часових відносин записами з камер. Зовнішній вигляд однієї й тієї ж людини може значно змінитися завдяки іншим об'єктам, таким як сумки, верхній одяг, вигляд зі спини тощо. Тому в межах класу варіації можуть бути великими, в той час міжкласові варіації можуть бути відносно меншими.

Навіть якщо дескриптори людей можуть бути ефективно знайдені, їх узгодження між камерами за наявності великої кількості спостережуваних людей є нетривіальною задачею. Порівняння дескрипторів осіб у великій кількості потенційних кандидатів є важким завданням, оскільки дескриптори фіксуються в різних місцях, моментах часу і для різної тривалістю. Складність процесу узгодження додатково збільшується, оскільки збільшення кількості кандидатів призводить до втрати специфіки дескриптора, збільшуючи можливість помилок при перевірці відповідності.

## **1.2 Загальний алгоритм повторної ідентифікації особи**

Система Re-ID є системою розпізнавання, схематично показаної на рисунку 1.1, яка складається з галереї (набір відомих екземплярів, Gallery set) і примірника (зображення особи, якому необхідно надати ідентифікатор, Probe ID).

Система Re-ID на початку обробки може мати:

- відкритий набір Re-ID дескрипторів — галерея може не містити відповідний ідентифікатор для запиту, тобто на вхід можуть прийти зображення людей,

ідентифікаторів яких ще немає в базі і ознаки яких необхідно додати до галереї.

- закритий набір Re-ID дескрипторів (с обмеженою формою відкритого набору) — коли можливі ідентифікатори запитів є підмножиною наявних у галереї ідентифікаторів, виконується порівняння зображення запиту з галереєю і обирається ідентифікатор особи з найвищим коефіцієнтом схожості із запитом.

Нехай набір відомих екземплярів буде представлений у вигляді  $G = (g_1, g_2, \dots, g_N)$ . Набір відомих ідентифікаторів задається  $id(G) = (id(g_1), id(g_2), \dots, id(g_N))$ , де функція  $id(.)$  вказує значення ідентифікатора, призначеного його аргументу. Нехай  $P = (p_1, p_2, \dots, p_M)$  являє собою набір примірників, яких ще немає в системі, їм відповідають ідентифікатори  $id(P) = (id(p_1), id(p_2), \dots, id(p_M))$ . Примірник подається в систему, де порівнюється з відомими екземплярами, при цьому обчислюються коефіцієнти подібності. Галерея ранжується за отриманими значеннями коефіцієнта подібності, для того щоб визначити ідентифікатор примірника.

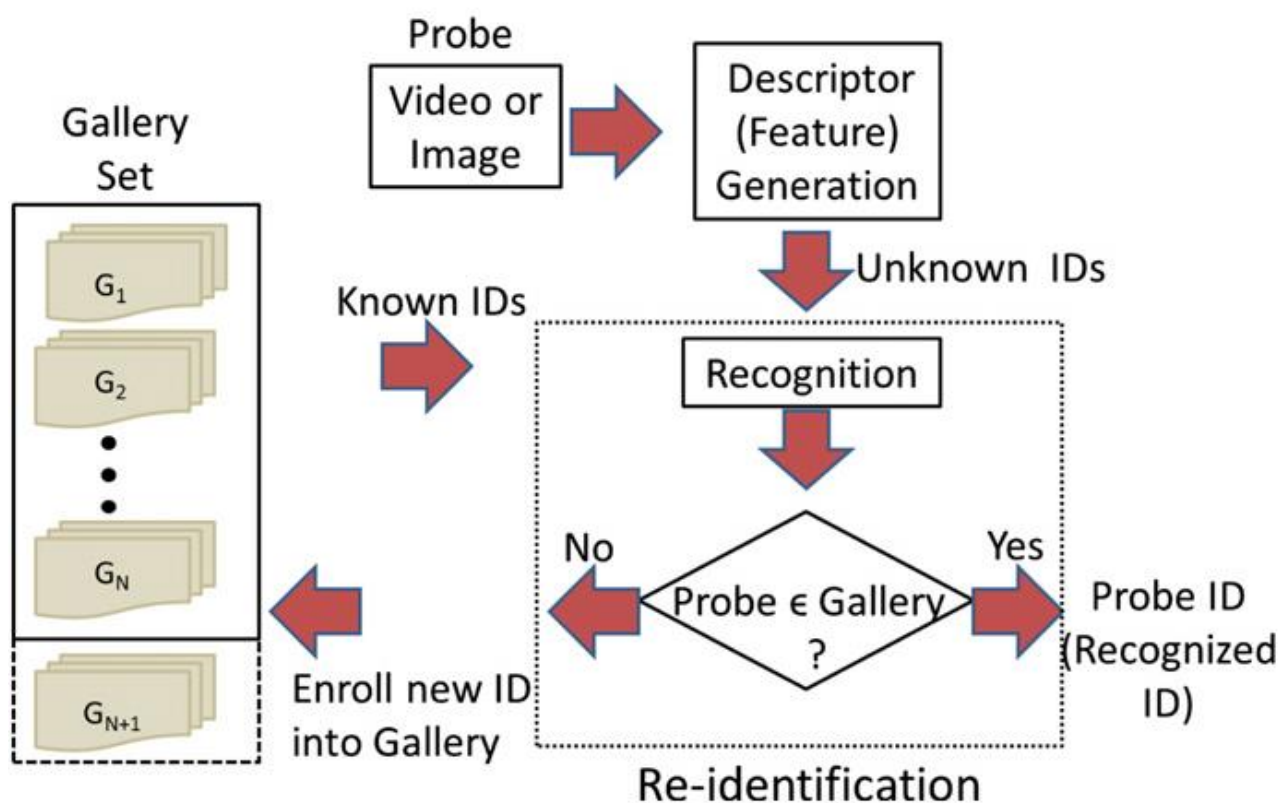


Рисунок 1.1 — Re-ID як розпізнавальна система [1]



Для закритого набору Re-ID, де ідентифікатор примірника існує в галереї,  $id(P) \subseteq id(G)$ , істинним ідентифікатором для даного примірника  $p_j \in id(p_j) = id(g_i^*)$ , таким чином

$$i^* = \operatorname{argmax}_{i \in 1, \dots, N} p(g_i | p_j) \quad (1.1)$$

де  $p(g_i | p_j)$  - ймовірність того, що  $id(p_j) = id(g_i)$ ,  $p$  найчастіше представлена коефіцієнтом подібності. Ідентифікатор екземпляру верхнього рангу присвоюється примірнику. З іншого боку, у відкритому наборі Re-ID примірник може бути або не бути підмножиною галереї. Основна задача при роботі з відкритим набором Re-ID полягає в тому, щоб спочатку встановити, чи є ідентифікатор примірника частиною галереї, і якщо так, визначити наявний ідентифікатор примірника. Для того, щоб знайти істинний ідентифікатор, крім ранжування елементів галереї і визначення  $i^*$  за допомогою формули (1.1), також має бути задоволена наступна умова:

$$p(g_i | p_j) > \tau \quad (1.2)$$

У формулі (1.2)  $\tau$  — вибраний поріг, при якому приймається, що  $id(P) \subseteq id(G)$ . Якщо ця умова не виконується, то визначається, що примірник не є частиною галереї. В такому випадку ідентифікатор примірника буде зареєстровано в галереї.

## Висновки до розділу 1

Повторна ідентифікація особи є завданням асоціювання зображень однієї особи з різних камер або з однієї камери в різних випадках. Це широка і складна задача з численними відкритими проблемами. Вона має три основні етапи: виявлення особи, відслідковування особи, визначення ідентифікатора особи, кожен з яких має свої виклики. Особи, які відслідковуються, можуть бути завчасно визначені та занесені в базу для подальшого порівняння з даними, які отримуються під час обробки відео, або заноситись в базу вже в процесі обробки, що є більш складним сценарієм.

Повторна ідентифікація вимагає надійні і унікальні дескриптори, автоматизація отримання яких є ресурсомісткою задачею, яка має багато перешкод через велику варіативність можливих положень людини та її зовнішнього вигляду.



## 2 ОГЛЯД ПІДХОДІВ ВИЯВЛЕННЯ І ВІДСТЕЖЕННЯ ЛЮДЕЙ НА ВІДЕО

### 2.1 Виявлення об'єктів

Виявлення об'єктів — завдання визначення присутності, розташування і протяжності будь-яких примірників об'єктів певного класу чи декількох класів, які присутні на зображенні. На зображенні може бути позначено обмежувальними прямокутними контурами або піксельними масками [4].

Виявлення людей є одним з перших завдань комп'ютерного зору, що почало застосовувати глибинне навчання.

Процес виявлення об'єктів складається з двох етапів:

- Визначення наявності об'єкта на зображенні (класифікація)
- Визначення розташування цього об'єкта (локалізація)

Основними проблемами для цієї задачі в контексті роботи з матеріалами відеоспостереження є:

- Пошук об'єктів невеликого розміру
- Накладання об'єктів, висока щільність об'єктів
- Наявність некоректних виявлень
- Програмно-апаратні обмеження (необхідність роботи в реальному часі)

#### 2.1.1 Методи виявлення об'єктів

Прогрес у виявленні об'єктів за останні два десятиліття пройшов через два історичні періоди:

- період виявлення класичними методами (до 2014 року)
- період виявлення на основі глибинного навчання (після 2014 року)

На рисунку 2.1 відображені основні етапи розвитку методів виявлення об'єктів:

## Object Detection Milestones

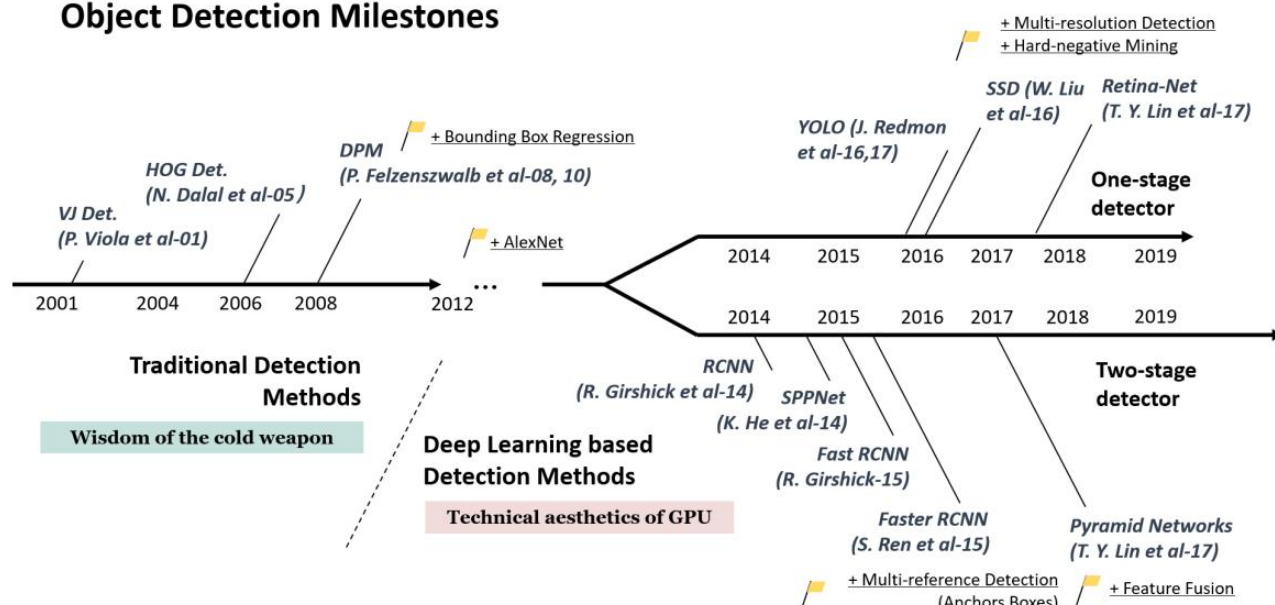


Рисунок 2.1 – Основні етапи розвитку методів виявлення об'єктів [5]

### Класичні підходи

Найбільш поширеними є Haar Cascade та Histogram of Oriented Gradients.

Haar Cascade — метод для виявлення об'єктів, який базується на ознаках Хаара, запропонований Р. Viola і М. Jones у своїй роботі «Швидке виявлення об'єктів з використанням підсиленого каскаду простих функцій», опублікованому в 2001 році. Цей підхід широко використовується для виявлення облич.

HOG (Histogram of Oriented Gradients) — підхід виявлення об'єктів, запропонований Н. Dalal та В. Triggs у своїй роботі «Гістограми орієнтованих градієнтів для виявлення людини», опубліковану в 2005 році.

### Недоліки ранніх підходів:

- Пропуски об'єктів

Класичні підходи мають проблеми у виявленні людей у різних позах, якщо для кожної пози використовується своя модель. Загально доступні попередньо навчені моделі призначені для виявлення прямостоячої людини. Вони добре виявляють осіб з лицьового та зворотнього боку. Проте виявлення осіб з бічних сторін в цілому є слабким.

- Помилкові виявлення та дуплікати

Об'єкти інших класів часто виявляються як людина. Досягти компромісу між пропущеними та помилковими виявленнями можна регулюючи значення параметрів алгоритмів. Деякі помилкові виявлення можна уникнути, визначивши значення мінімального розміру об'єкту виявлення.

- Ненадійна межа виявлення

Межа виявлення, яку знаходить Haar Cascade і HOG, не щільно обмежує виявлену людину. Межа цього контуру не узгоджується між виявленнями одного і того ж самого об'єкту на послідовних кадрах. Це ускладнює визначення положень частин тіла людини (наприклад, розташування ніг) з використанням співвідношень, розрахованих на межі виявлення.

- Пропуски об'єкту при виявленні на послідовних кадрах

Досить часто спостерігається, що людина, виявлена в одному кадрі, не виявляється в наступному кадрі і навпаки.

### **Сильні сторони ранніх підходів:**

Незважаючи на вказані недоліки, ці підходи досі використовуються в індустрії. Вони вимагають меншої обчислювальної потужності порівняно з сучасними підходами на основі глибинного навчання (не мають потреби в роботі GPU в режимі реального часу). Завдяки реалізації в бібліотеках комп'ютерного зору, таких як OpenCV ці алгоритми є легкими для застосування.

Більшість недоліків ранніх підходів виявлення людини були розв'язані в нових підходах, які ґрунтуються на глибинному навчанні.

### **Підходи, які використовують глибинне навчання**

Сучасні підходи до виявлення людини характеризуються такими особливостями:

- в значній мірі базуються на глибинних згорткових нейронних мережах (Deep Convolution Neural Network, CNN). CNN широко адаптовані до різних проблем комп'ютерного зору, таких як класифікація зображень (визначення того, який тип об'єкта містить зображення), виявлення об'єктів (знаходження різних типів об'єктів у зображенні) та локалізації об'єкта (визначення

місцезнаходження виявлених об'єктів). Виявлення людини, є особливим випадком виявлення об'єктів і локалізації об'єкта.

- є багатокласовими детекторами об'єктів, які можуть виявляти кілька типів об'єктів, включаючи людей.

### **Недоліки, порівняно з класичними підходами**

Загалом, глибинні нейронні мережі забезпечують більш точні результати ціною більшої кількості обчислень. Моделі, які не використовують графічні прискорювачі мають невисоку точність або витрачають занадто багато часу на обробку одного зображення. Проте за наявності графічних прискорювачів, сучасні методи глибинного навчання здатні забезпечити ці поліпшені результати з порівняними частотами кадрів.

### **2.1.2 Метрика оцінки виявлення об'єктів**

Оцінка якості детектування має одночасно враховувати і класифікацію, і локалізацію. Метрикою, яка одночасно може оцінити ці задачі, є Average Precision (AP) [6],

$$AP = \frac{1}{I} \sum_{Recall_i} Precision(Recall_i) \quad (2.1)$$

яка пов'язана із визначенням помилок першого і другого роду, де Precision – точність, а Recall – повнота

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

Для них важливими є такі компоненти:

TP (True Positive) – кількість об'єктів, які належать до позитивного класу і класифіковані як позитивні.

FP (False Positive) – кількість об'єктів, які належать до негативного класу і класифіковані як позитивні.

FN (False Negative) – кількість об'єктів, які належать до позитивного класу і класифіковані як негативні.

$$Recall_i = [0, 0.1, 0.2, \dots, 1.0]$$

$Precision(Recall_i)$  – значення для Precision на кривій ROC-AUC (рисунок 2.2) у точках, які приймають значення із  $Recall_i$

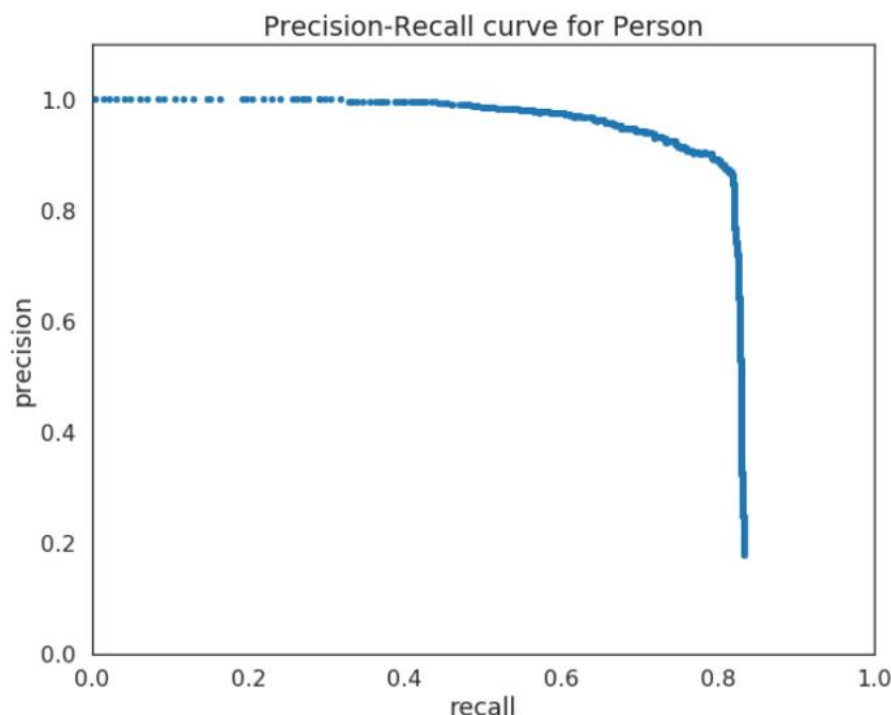


Рисунок 2.2 – Приклад побудови кривої ROC-AUC [7]

Для оцінки локалізації використовується метрика Intersection over Union (на основі коефіцієнта Жаккара)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2.4)$$

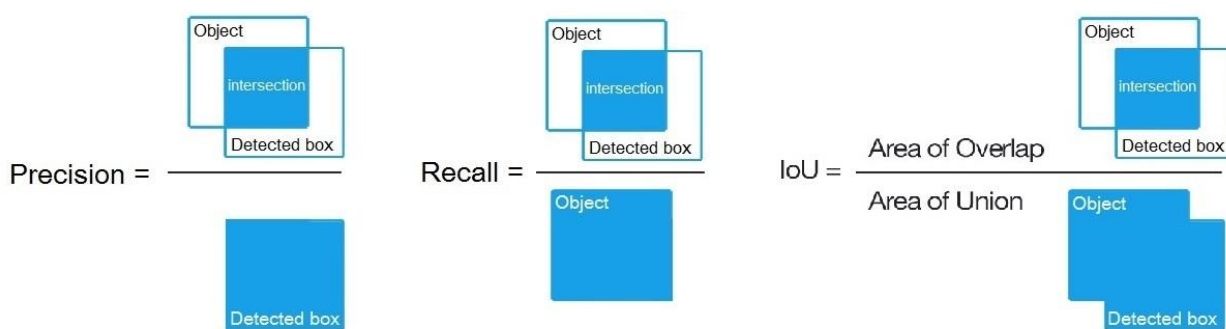


Рисунок 2.3 – Візуалізація знаходження коефіцієнту IoU [8]

Ці дві метрики поєднуються завдяки введенням таких означень для TP, FP, FN:

TP – кількість детекцій з  $IoU > 0.5$

FP – кількість детекцій з  $IoU \leq 0.5$  або знайдені більше одного разу

FN – кількість об'єктів, які не були знайдені, або були знайдені з  $IoU \leq 0.5$

Для оцінки моделей, які виявляють декілька класів, використовується метрика mAP (mean Average Precision). Є декілька протоколів її підрахунку, проте в загальному вигляді mAP має такий вигляд:

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (2.5)$$

де Q це кількість наявних у тренувальному наборі класів. Чим вищий показник mAP, тим точніші виявлені рамки об'єктів.

## 2.2 Відстеження декількох об'єктів

Multi Object Tracking (MOT) — має на меті оцінку траєкторій декількох об'єктів, знаходячи їх розташування та зберігаючи значення ідентифікаторів на наступних кадрах відео. Існуючі методи MOT можуть бути поділені на офлайн і онлайн. Офлайн методи MOT використовують для генерації траєкторій як попередні, так і наступні кадри, в той час як онлайн методи MOT використовують тільки інформацію, доступну до поточного кадру. Хоча в офлайн методах є певні переваги при обробці неоднозначних результатів відстеження, вони не можуть бути застосовані в режимі реального часу [9].

Алгоритми відстеження є швидшими за алгоритми виявлення об'єктів, оскільки вони виконують локальний пошук замість глобального пошуку. Завдяки цьому ми можемо отримати дуже високу частоту кадрів для нашої системи, виконуючи виявлення об'єктів на кожному n-му кадрі і відстежуючи об'єкт у проміжних кадрах.

На деяких кадрах виявлення об'єктів може бути невдалим, проте відстеження об'єктів все ще може бути можливим, оскільки відстеження враховує місцерозташування і зовнішній вигляд об'єкта у послідовності попередніх кадрів.

Алгоритм відстеження іноді може втратити об'єкт, який він відстежує. Наприклад, коли рух об'єкта занадто швидкий, або частина об'єкта була закрита іншим об'єктом. Тому для побудови відстежувальних систем реальні застосунки використовують комбінацію методів виявлення та відстеження.



Останні методи MOT переважно приймають стратегію відстеження за виявленням і виконують завдання пов'язуючи виявлення між кадрами за допомогою алгоритмів асоціювання даних, тому ці підходи значною мірою покладаються на якість результатів виявлення. Якщо виявлення відсутнє або неточне, цільовий об'єкт швидше за все буде втрачено. Для вирішення таких проблем останні запропоновані методи використовують методи відстеження кожного об'єкта для MOT. Трекер кожного об'єкта використовує виявлення в першому кадрі і оновлює модель, щоб знайти об'єкт в наступних кадрах. Проте, він схильний до втрати об'єкта, коли ціль перекривається іншими об'єктами. Асоціювання даних застосовується, коли оцінка відстеження знаходиться нижче порогового значення, що вказує, що цільовий об'єкт втрачений або зазнав суттєвих змін у зовнішньому вигляді. Основний виклик використання власного трекера для кожного об'єкта в MOT полягає в тому, щоб впоратися з частими взаємодіями між об'єктами і міжкласовими дистракторами. Існуючі методи відстеження одного об'єкта знаходяться під впливом проблеми дисбалансу даних між позитивними та негативними зразками для оновлення моделей. У області пошуку трекера лише декілька місць поблизу центру шуканого об'єкта відповідають позитивним зразкам, тоді як всі зразки, взяті на інших позиціях, є негативними. Більшість місць з фонові області є негативами, що може призвести до неефективного тренування моделі та послабити її дискримінаційну силу. Ця проблема загострюється в контексті завдання MOT. Якщо модель переповнена фоновими негативами, трекер схильний до дрейфу, коли подібні дистрактори з'являються в області пошуку. Таким чином, необхідно зосередити увагу на невеликій кількості прикладів під час оновлення моделі, щоб полегшити проблему дрейфу пошукового вікна. Для асоціації даних необхідно порівняти поточну виявлену ціль з послідовністю попередніх спостережень у траєкторії. Неточні виявлення та виявлення з перекриттям в попередніх спостереженнях траєкторії призводять до шумних оновлень і роблять модель зовнішнього вигляду менш ефективною. На рисунку 2.4 зображено основні етапи MOT на прикладі відслідковування одного об'єкта.

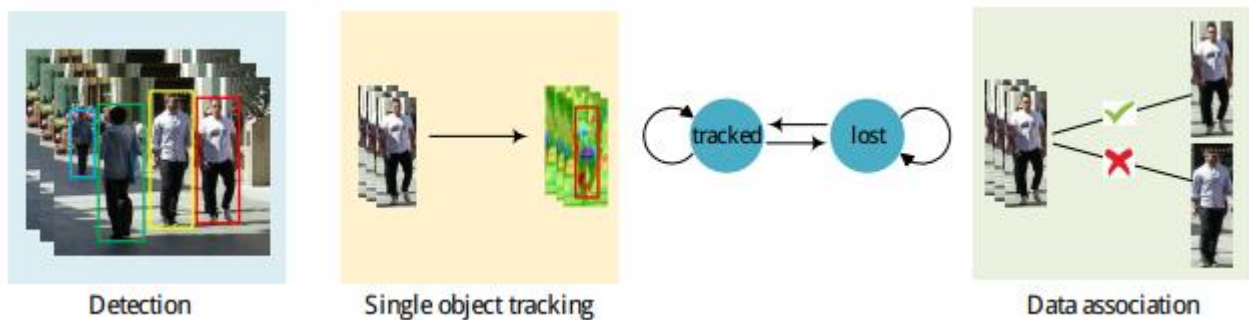


Рисунок 2.4 — Основні етапи MOT [9]

### 2.2.1 Формалізація проблеми відстеження декількох об'єктів

Загалом, відстеження декількох об'єктів можна розглядати як проблему багатоваріантної оцінки. Дано послідовність зображень,  $s_t^i$  позначення стану  $i$ -го об'єкта в  $t$ -му кадрі,  $S_t = (s_t^1, s_t^2, \dots, s_t^{M_t})$  — стани всіх  $M_t$  об'єктів в  $t$ -му кадрі.  $s_{i_s:i_e}^i = \{s_{i_s}^i, \dots, s_{i_e}^i\}$  — послідовні стани  $i$ -го об'єкта, де  $i_s$  та  $i_e$  є відповідно першим і останнім кадром, в якому наявний об'єкт  $i$ ,  $S_{1:t} = \{S_1, S_2, \dots, S_t\}$  — позначення всіх послідовних станів об'єктів від першого кадру до  $t$ -го кадру. Номер об'єкта може змінюватися від кадру до кадру. Згідно з найбільш часто використовуваної парадигмою відстеження за виявленням (Detection Based Tracking, DBT),  $o_t^i$  використовується для позначення зібраних спостережень для  $i$ -го об'єкта в  $t$ -му кадрі,  $O_t = (o_t^1, o_t^2, \dots, o_t^{M_t})$  для позначення зібраних спостережень для всіх  $M_t$  об'єктів у  $t$ -му кадрі, а  $O_{1:t} = \{O_1, O_2, \dots, O_t\}$  для позначення всіх зібраних послідовних спостережень всіх об'єктів від першого кадру до  $t$ -го кадру. Метою MOT є пошук «оптимальних» послідовних станів всіх об'єктів, які можна взагалі моделювати, виконуючи MAP (maximal a posteriori, максимальну апостеріорну) оцінку з умовного розподілу послідовних станів з урахуванням всіх спостережень:

$$\hat{S}_{1:t} = \operatorname{argmax}_{S_{1:t}} P(S_{1:t} | O_{1:t}) \quad (2.6)$$

### 2.2.2 Огляд класичних методів відстеження об'єкта

- Присвоєння ідентифікаторів на основі центроїд

у найпростішій формі ідентифікатори можуть бути призначені завдяки центроїдам обмежувального контуру. Спочатку обчислюються центроїди для кожного контуру у кадрі 1, у кадрі 2 розглядаються нові центроїди і на основі відносної відстані від попередніх центроїд їм призначаються ідентифікатори. Основне припущення полягає в тому, що контур об'єкта для підрахунку центроїд буде лише трохи рухатися. Цей простий підхід працює досить добре, поки центроїди знаходяться далеко одна від одно та виходить з ладу, коли люди знаходяться близько один до одного, оскільки в результаті буде відбуватись часте присвоєння хибних ідентифікаторів.

#### - Фільтр Калмана

Поліпшення порівняно з простим відстеженням на основі центроїд. Фільтр Калмана дозволяє відстежувати модель на основі положення і швидкості об'єкта та прогнозувати його імовірне положення, моделюючи майбутнє положення і швидкість за допомогою гаусіанів. Коли він отримує на вхід новий примірник, він може використовувати ймовірність для призначення свого прогнозу і оновлення. Фільтр Калмана потребує менше пам'яті і є більш швидким. Оскільки він використовує як положення, так і швидкість руху, то показує кращі результати, ніж відстеження на основі центроїд.

## **Висновки до розділу 2**

В цьому розділі було розглянуто загальні принципи виявлення та відслідковування об'єктів, їх теоретичні основи та історичну обумовленість. Виявлення об'єктів є задачею класифікації та локалізації, для якої існує багато різноманітних підходів. Було описано метрику AP, яка дозволяє оцінити якість детектуючого алгоритму.

Для задачі реідентифікації необхідним є відстеження декількох об'єктів одночасно, це досягається використанням алгоритмів MOT, для яких було описано основні виклики і проблеми та математично визначено проблему відслідковування об'єктів і описано базові підходи.



## 3 ВИБІР АРХІТЕКТУР НЕОБХІДНИХ НЕЙРОННИХ МЕРЕЖ

### 3.1 Глибинні нейронні мережі

Глибинне навчання — це метод машинного навчання, який дозволяє навчати модель передбачати результат по набору вхідних даних. Для навчання мережі можна використовувати як навчання з учителем (supervised learning), тобто з наявним тренувальним анотованим набором даних, так і навчання без вчителя (unsupervised learning). Існує три типи шарів нейронів в нейронній мережі: вхідний шар, прихований шар (шари) і вихідний шар. Зв'язки між нейронами мають вагу, який визначається важливістю елемента вхідних даних. Глибинною нейронною мережею називається мережа, яка має більше двох прихованих шарів. Для навчання глибокої нейронної мережі необхідно мати великий набір даних.

Вхідний шар приймає вхідні дані і передає ці дані в перший прихований шар. Приховані шари виконують математичні обчислення з вхідними даними. Одне із завдань при проектуванні архітектури нейронних мереж — визначення кількості прихованих шарів і нейронів на кожному шарі. Вихідний шар видає результат. На рисунку 3.1 схематично зображено загальну модель нейронної мережі.

Для навчання глибинної мережі необхідні:

- великий набір даних.
- великі обчислювальні потужності.

Ітераційно порівнюючи вихідні результати з включеними в набір даними, можна обчислити функцію втрат, що вказує, наскільки сильно помиляється алгоритм. Після кожної ітерації (epoch) ваги між нейронами перерозподіляються за допомогою методу градієнтного спуску для мінімізації функції втрат.

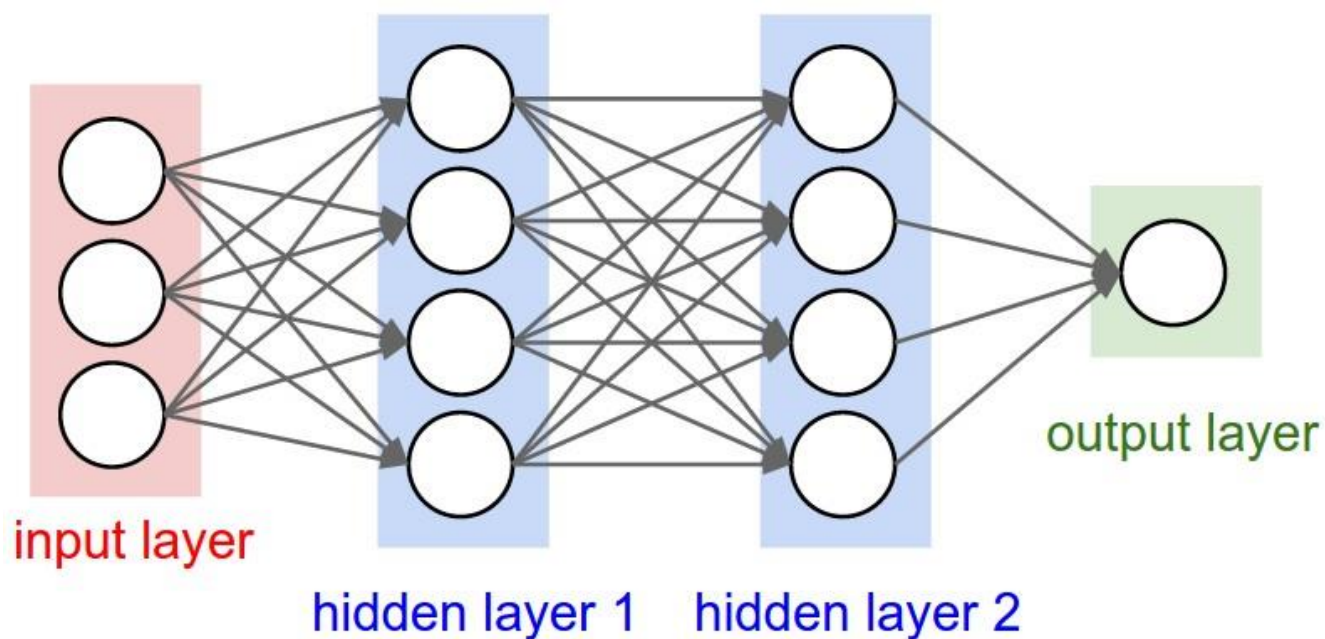


Рисунок 3.1 — Схематичне зображення нейронної мережі [10]

Використання нейронних мереж можливе у декількох варіантах:

- Тренування штучної нейронної мережі на власних даних
- Використання обраної архітектури штучної нейронної мережі з готовими вагами (pretrained model), натренованої на певному наборі даних для певної кількості класів
- Дотреновування штучної нейронної мережі своїми даними (fine-tuning, transfer learning)

Для тренування нейронної мережі на власних даних вони мають бути анотовані, тобто мати значення правильних міток (ground truth values), які має передбачати нейронна мережа. Отримання таких анотацій зазвичай потребує значних ресурсів. Модель, натренована на невеликому наборі даних, матиме тенденцію до перенавчання, тобто показувати слабкі результати для примірників, які не входили в тренувальні дані, і матиме слабку генералізацію на зміну умов отримання зображення. Тому набір даних для тренування має мати якомога більше різноманітних варіацій (освітлення, кут повороту, розмір, колір, форма тощо). Самостійне тренування моделі вимагає значних обчислювальних потужностей, проте дозволяє обмежити класи лише до необхідних для задачі. Найоптимальнішим варіантом є дотреновування готової моделі з використанням своїх даних.

Глибинні нейронні мережі навчаються з нуля за допомогою величезних наборів даних, які містять мільйони зображень і зазвичай добре узагальнюються для великої кількості класів. Хоча з усіх цих класів для задач переважно використовується лише декілька, багато зображень об'єктів різних класів мають спільні риси, а функції одного детектора об'єктів одного класу добре працюють під час пошуку об'єкта іншого класу, що дає змогу перетренувати кілька останніх шарів, приймаючи решту як вже побудований екстрактор ознак.

### **3.2 Вибір архітектури штучної нейронної мережі для виявлення людей**

Існує безліч алгоритмів для виявлення об'єктів певного класу на фото або відео. Останні п'ять років спостерігався стрімкий розвиток згорткових нейронних мереж (CNN). Їх архітектура дозволила зробити модель детектування, яка самостійно знаходить в зображенні високрівневі ознаки, використовуючи лише зображення у якості вхідних даних.

Сучасні детектори можна умовно поділити на одностадійні та двостадійні. Домінуючою парадигмою в сучасному виявленні об'єктів є двоступеневий підхід. Перший етап генерує розріджений набір можливих розташувань об'єктів, а другий етап класифікує кожного знайденого кандидата як один із заданих класів або як частину фону з використанням згорткової нейронної мережі.

Одностадійні детектори застосовуються на регулярній, щільній вибірці місць розташування об'єктів, масштабів і пропорцій і оброблюють все зображення за один прохід алгоритму. Останні алгоритми одностадійних детекторів, таких як YOLO і SSD, демонструють багатообіцяючі результати з точністю в межах 10-40% відносно сучасних двостадійних методів [11]. Ці детектори орієнтовані на швидкість. Останні роботи показали, що час роботи двостадійних детекторів може бути зменшено за рахунок зменшення роздільної здатності вхідного зображення, і мають вищу точність навіть при більшій обчислювальній потужності для одностадійних методів.

Одним із найоптимальнішим сучасним детектором відносно часу роботи та точності є YOLOv3 (You Only Look Once version 3 [12]). Перевага цього методу

полягає в тому, що цей детектор є швидшим, ніж загальноприйняті двоетапні детектори, наприклад Faster R-CNN. YOLO об'єднує дві стадії в одну нейронну мережу. Проста архітектура робить можливим виявлення об'єктів в реальному часі (проте з використанням потужних графічних прискорювачів).[13]

### **3.3 Вибір архітектури штучної нейронної мережі для відстеження людей**

Одним із найбільш ефективних алгоритмів для відстеження людей є Simple online and realtime tracking (SORT[14]), який застосовує фільтр Калмана у просторі зображень та асоціює дані кадр за кадром, використовуючи угорський метод з метрикою асоціації, яка вимірює перекриття обмежувальних рамок. Цей підхід працює з високою частотою кадрів і в той же має високі якісні показники. Досягаючи загальної продуктивності з точки зору метрик Precision та Accuracy, SORT часто повертає нові індикатори для вже визначених об'єктів. Це відбувається, тому що використовувана метрика асоціації є точною лише тоді, коли невизначеність оцінки стану є низькою. SORT має вади у відстеженні через накладання об'єктів, які з'являються в камерах з фронтальним переглядом.

Шлях до розв'язку цієї проблеми пропонує алгоритм DeepSORT[15], метрика асоціації у якому поєднує інформацію про рух і зовнішній вигляд. Зокрема застосовується згорткова нейронна мережа (CNN), яка була підготовлена для виділяти пішоходів на великому наборі даних для повторної ідентифікації людини. Завдяки інтеграції цієї мережі цей алгоритм підвищує надійність обробки пропусків виявлень і накладань об'єктів, зберігаючи при цьому систему легкість реалізації, ефективної і застосовності до роботи в реальному часі. Код і попередньо підготовлена модель CNN є загальнодоступними для полегшення експериментальних досліджень і розробки практичних застосунків.

### **3.4 Застосування нейронних мереж для реідентифікації**

Сучасні алгоритми відслідковування об'єктів зберігають ідентифікатор об'єкта поки він знаходиться в полі зору камери, проте при новій його появі черей певний час, йому буде присвоєно новий ідентифікатор. Цю проблему може бути



поделано завдяки використанню алгоритму реідентифікації. Після відслідковування отримуються дані про послідовності обмежувальних контурів з однаковими ідентифікаторами, ці послідовності можуть бути передані на вхід нейронної мережі, яка витягає з них унікальний вектор ознак. Роботу такої мережі схематично зображено на рисунку 3.2.

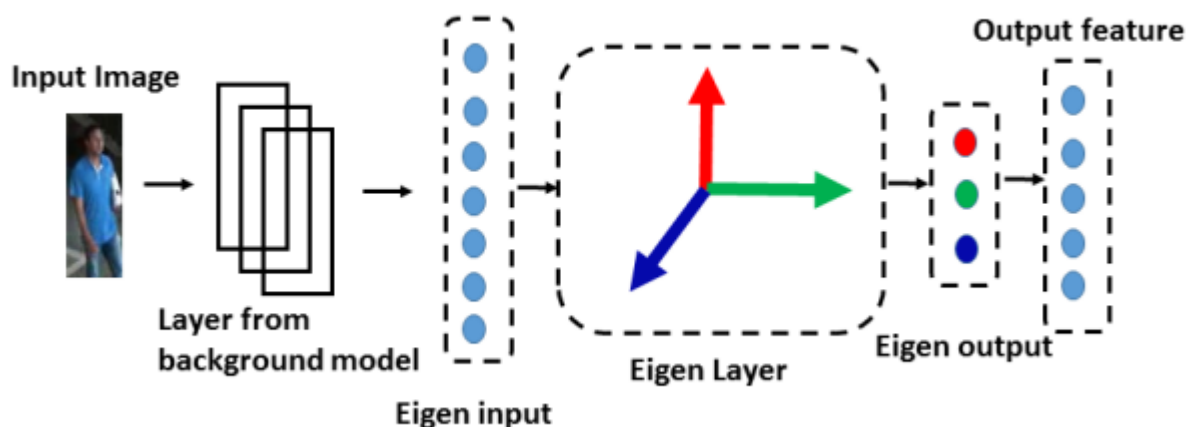


Рисунок 3.2 — Отримання вектора ознак для зображення [1]

Знайдені ознаки записуються в базу, надалі з наступних знайдених алгоритмом відслідковування послідовностей знову витягуються вектори ознак і порівнюються з вже занесеними в базу (рисунок 3.3), отримані значення ранжуються, якщо для першого рангу значення метрики асоціації вище визначеного порогу, примірнику назначають відповідний ідентифікатор. В іншому випадку ознаки послідовності вносяться в базу.

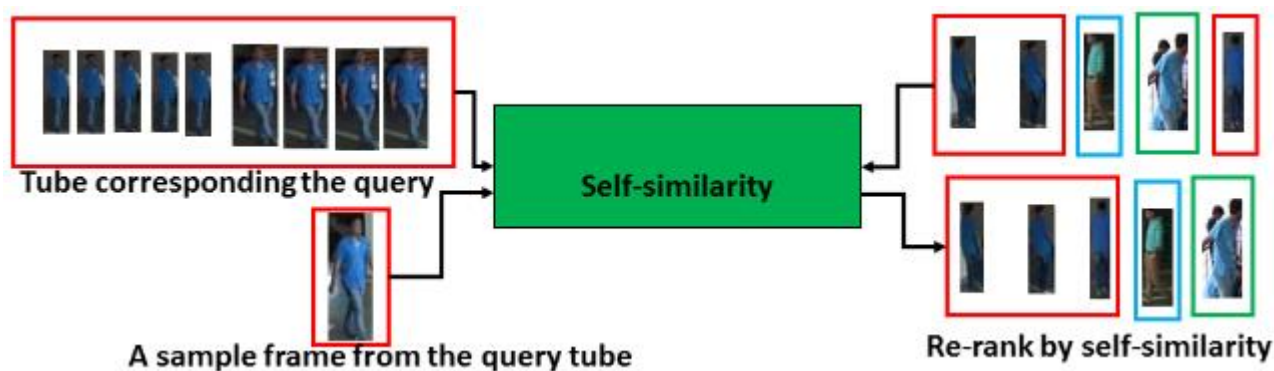


Рисунок 3.3 — Порівняння запиту з існуючою в базі послідовністю [16]

### **Висновки до розділу 3**

У розділі 3 було розглянуто основні теоретичні положення нейронних мереж та переваги їх застосування для задач виявлення та відслідковування об'єктів. Було розглянуто методи виявлення і відстеження людей в контексті вибору підходів на основі глибинного навчання для їх практичного застосування до відеоматеріалу. Значну увагу було приділено одностадійним детекторам YOLO та SSD, які є компромісом між швидкістю та точністю серед моделей нейронних мереж.

Було описано принципи застосування нейронних мереж для реідентифікації на основі інформації, яка отримується після використання алгоритмів виявлення та відслідковування об'єктів.

## 4 РОЗРОБЛЕННЯ ПРОГРАМНОЇ РЕАЛІЗАЦІЇ

### 4.1 Опис вхідних даних

Вхідними даними є 10 відео, які були послідовно зняті однією камерою відеоспостереження.

Основні характеристики відеоматеріалу:

- Загальна довжина: 3:31:38
- Роздільна здатність: 1280 x 720 пікселів
- Частота кадрів: 20 fps (всього 253960 кадрів)

Проблеми в даних, важливі для застосування алгоритмів комп'ютерного зору:

- наявність сильно вираженого радіального викривлення через відсутність калібрування у камери. Викривлення візуально деформують силует людини і система має пристосовуватись до різних можливих деформацій. Якщо є фізичного доступ до камери, то калібрування можна проводити за допомогою шахової дошки [17]. Через радіальне викривлення на бічних сторонах результуючого зображення при виявленні одного об'єкту в його знайденому обмежувальному контурі повноцінно знаходиться ще мінімум один об'єкт (рисунок 4.1 а)
- кут огляду камери, який в даному відеоматеріалі сприяє створенню великої кількості накладань зображень людей, що ускладнює використання алгоритмів відслідковування
- на деяких відео занадто великий потік людей, який складно розрізнити навіть людині
- наявність зображень однієї і тієї самої людини в верхньому одязі і без, що ускладнює ідентифікацію однієї і тієї самої особи
- велика кількість людей в уніформі, що ускладнює виділення унікальних дескрипторів для персоналу (рисунок 4.1 б)
- розмиття об'єктів, які швидко рухаються (рисунок 4.1 в)



Рисунок 4.1 — Приклади особливостей вихідного відео

## 4.2 Застосування готових програмних рішень

Оскільки задача реідентифікації знаходиться на даній в активному науковому розвитку, існує багато створених підходів і алгоритмів, проте вони рідко реалізовані програмно. Intel OpenVINO [18] надає готовий програмний продукт, який використовує натреновані моделі детекції та реідентифікації, проте на вихідних даних це програмне забезпечення показує слабкі результати. Найкраще показали себе такі моделі:

- person-detection-retail-0013
- person-reidentification-retail-0031



Рисунок 4.2 — Приклад детекцій OpenVINO

### 4.3 Власна програмна реалізація

Для програмної реалізації використовувалась високорівнева мова Python, яка є однією з найпоширеніших мов програмування для аналізу зображень, для роботи з зображеннями і відео — бібліотека OpenCV, для застосування машинного навчання — бібліотека PyTorch. Задача має відкритий набір Re-ID дескрипторів, тобто немає завчасно визначених осіб для пошуку.

Розрахунки проводились з використанням графічного прискорювача Nvidia GeForce GTX 1060 3GB (Nvidia compute capability: 6.1), CUDA версії 10.0, PyTorch версії 1.1.

#### 4.3.1 Створення анотацій

Для покращення застосування алгоритмів виявлення людей і реідентифікації було створено тренувальні дані. Для цього було взято одне з відео і з нього було обрано 200 кадрів, на яких було вручну визначено координати об'єктів людей (x, y координати верхнього кута обмежувального прямокутника та висота і ширина цього прямокутника, які разом однозначно визначають обмежувальний контур).



Отримані дані про 100 кадрів було використано для дотреновування моделі виявлення людей, інші 100 були використані для оцінки метрик AP і Rank1.

Таблиця 4.1 — Опис анотованих даних

Індекси кадрів	Кількість об'єктів	Кількість унікальних осіб
0..99	914	92
100..199	893	89

До тренувальних даних було застосовано штучні перетворення, для збільшення варіативності зображень, такі як відображення по горизонталі, зміна яскравості та контрасту, додавання шуму і розмиття. Приклади застосованих перетворень відображено на рисунку 4.3.

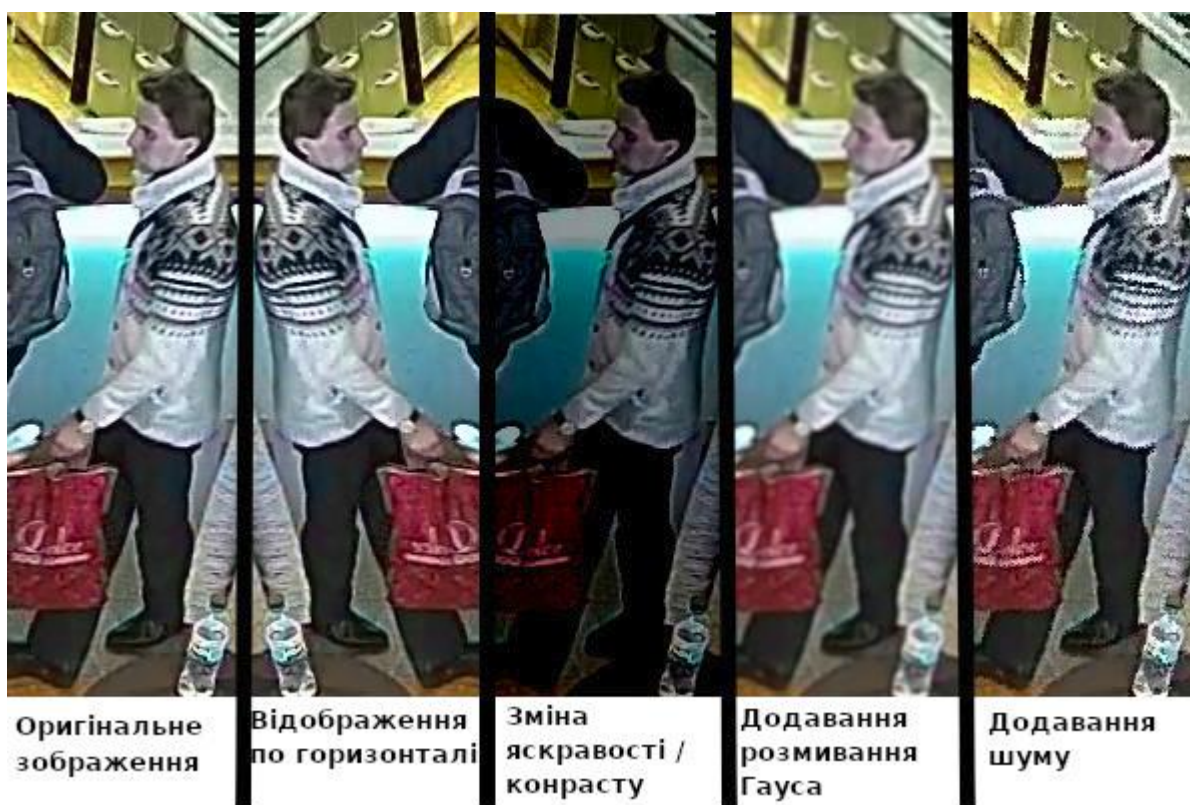


Рисунок 4.3 — Застосовані перетворення до тренувальних зображень

#### 4.3.2 Виявлення людей

Для виявлення людей була взята моделі YOLOv3 та Faster R-CNN ResNet-50 FPN, попередньо натреновані на COCO Dataset [19]. COCO Dataset складається з 123287 зображень, 886284 об'єктів та 80 класу. Для даної задачі використовується

лише клас “person”. Моделі були дотреновані тренувальними розміченими даними для більш точного детектування людей.

Піксель — найменша одиниця двовірного цифрового зображення, яка має певний колір, градацію сірого кольору і прозорість, а також форму. Подання растрового зображення в пам'яті комп'ютера - масив відомостей про колір всіх пікселів, упорядкований певним чином. Широкий діапазон кольорів можна вказати, використовуючи 3 байти - по одному на кожен колір Red, Green і Blue [20]. На вхід подавались зображення без зміни їх розміру, тобто масив розмірності 1280x720x3.

На рисунку 4.4 зображені приклади детекцій використаної моделі YOLOv3, середня частота обробки складає 1.5 fps. Двостадійна модель Faster R-CNN ResNet-50 FPN більш точно обмежує людей, проте має зайві детекції об'єктів, які не є людьми, приклад виявлень зображено на рисунку 4.5. Їх можна врегулювати шляхом встановлення додаткового порогу впевненості детекцій нейронної мережі.

Завдяки поєднанню результатів моделей можна досягти більшої точності детекцій, тому було створено ансамбль цих моделей.



Рисунок 4.4 – Приклади детекцій YOLOv3



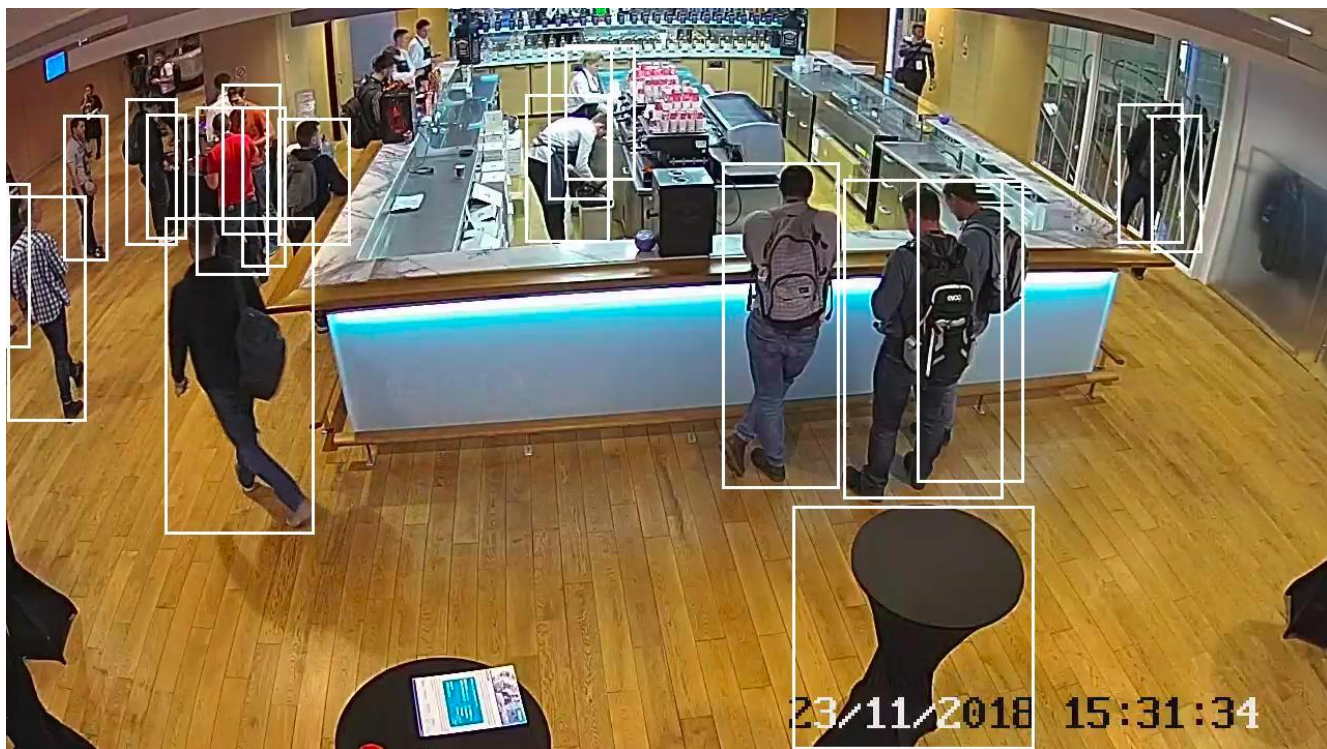


Рисунок 4.5 – Приклади детекції Faster R-CNN ResNet-50 FPN

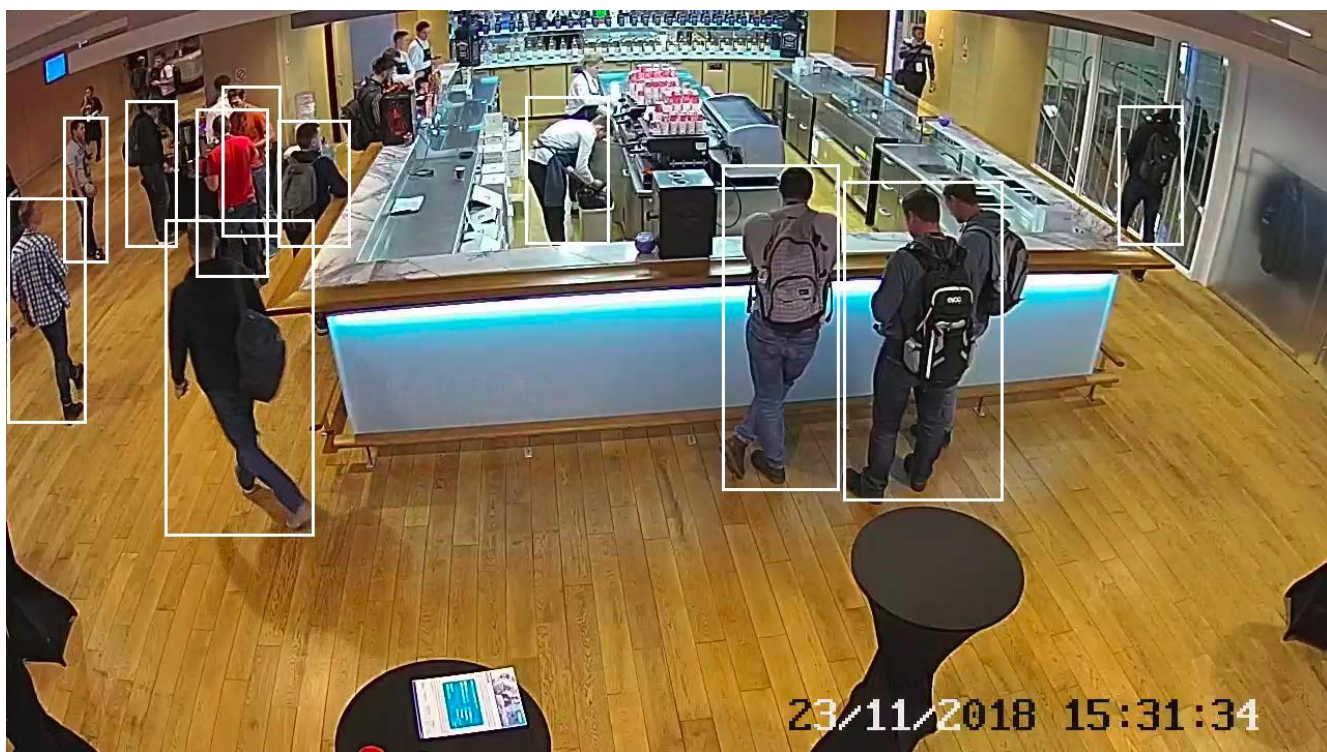


Рисунок 4.6 — Ансамбль моделей

Таблиця 4.2 — Значення метрики Average Precision на тестових даних

Модель	AP
--------	----



YOLOv3	63.8
Faster R-CNN ResNet-50 FPN	65.1
Ансамбль моделей	68.3

### 4.3.3 Відслідковування людей

Після знаходження детекцій на кадрі, вони оброблювались алгоритмом Deep SORT. Обробка зображень детектором і трекером відбувається з частотою кадрів 0.5fps, що недостатньо для роботи в реальному часі, проте уможлиблює її при використанні декількох більш потужних графічних прискорювачів та обробці кожного n-го кадру



Рисунок 4.7 — Приклад відпрацювання алгоритму Deep SORT

В результаті кожному відслідкованому об'єкту було надано певний ідентифікатор та отримано послідовності зображень для кожного такого ідентифікатора.



#### 4.3.4 Застосування реідентифікації

Всі виявлені і відслідковані примірники було виділено і збережено окремо для подальшого аналізу. За знайденими послідовностями зображень було натреновано нейронну мережу для реідентифікації, на рисунку 4.8 зображено приклади використаних послідовностей.

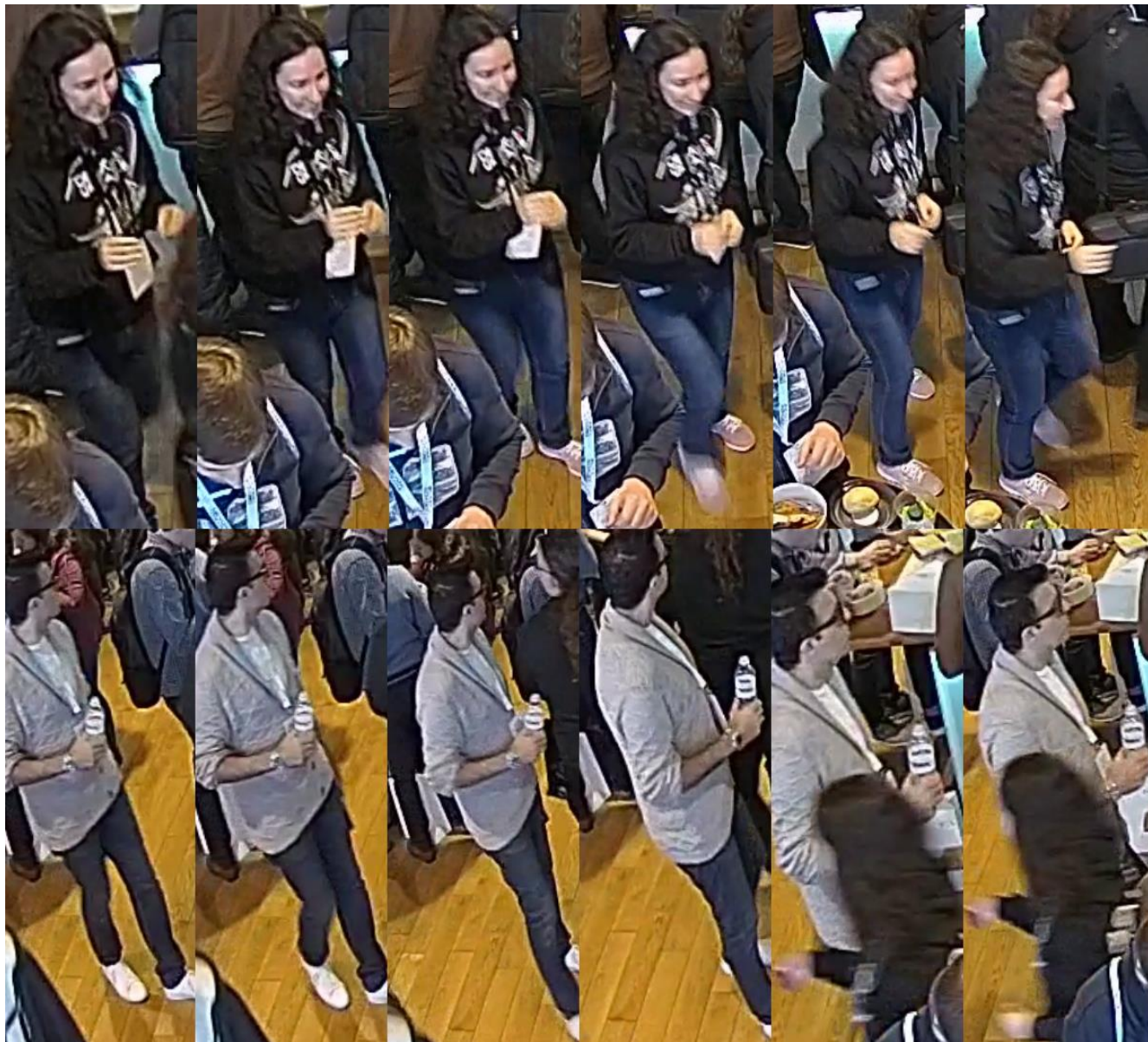


Рисунок 4.8 — Приклади об’єктів з однаковими ідентифікаторами

Значення метрики Rank1, яка показує відсоток правильно визначених ідентифікаторів, на тестових даних складає 86.2.

Після відпрацювання моделі реідентифікації її результати можна використовувати для побудови траєкторії руху об’єктів та визначенню кількості їх появи в кадрі за певний час. На рисунку 4.9 показаний приклад побудови такої

траєкторії, були взяті обмежувальні контури з однаковими ідентифікаторами та обраховано центри цих прямокутників, за цими точками за допомогою та згладженню ліній та кольорового градієнта по часу було візуалізовано шлях обраного об'єкта.

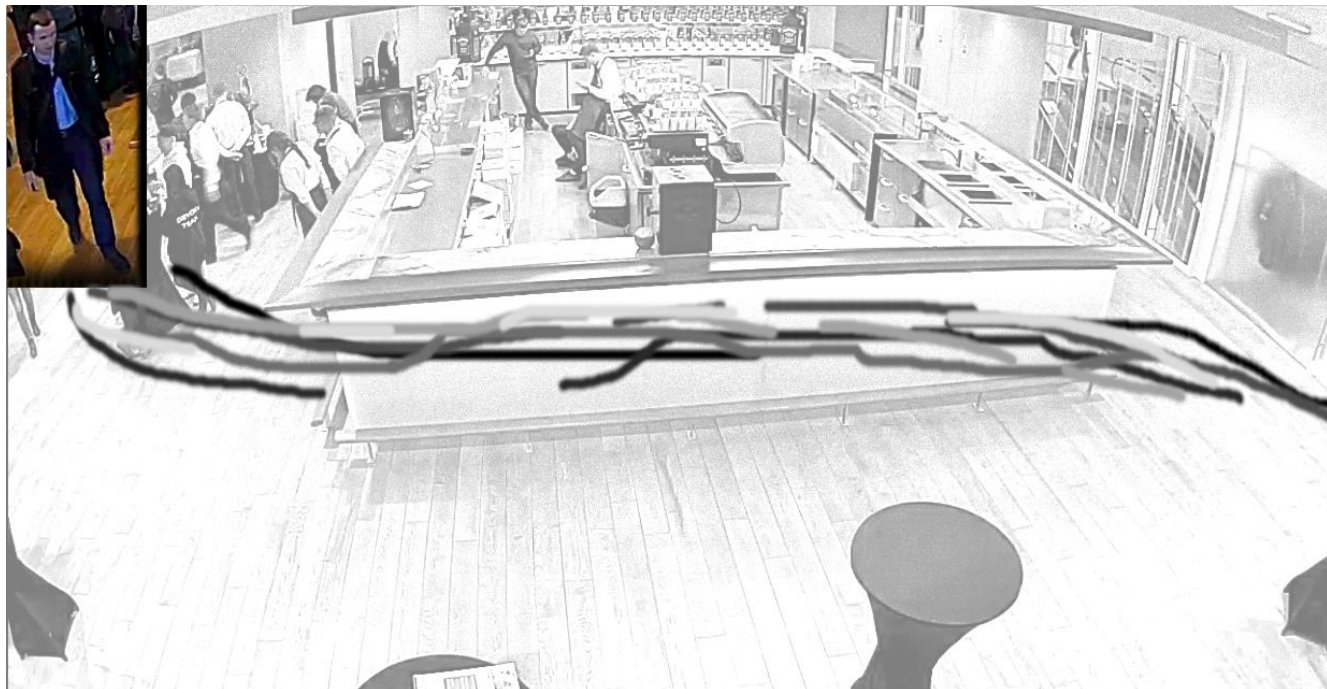


Рисунок 4.9 — Траєкторія руху певного вибраного об'єкту

#### 4.4 Аналіз отриманих результатів

Результати роботи алгоритму на частині проанатованих тестових даних відображено у таблиці 4.3

Таблиця 4.3 — Оцінка роботи системи на анованих тестових даних

	Кількість унікальних об'єктів
Реальне значення	89
Створена система	147



Найкраще обрана реалізація працює на статичних об'єктах, для рухомих об'єктів часто відбувається втрата ідентифікатора та присвоєння того ж самого ідентифікатора іншому об'єкту через його перекриття іншими об'єктами.

Всього було виявлено 731 унікальний об'єкт.

Завдяки різниці між номерами кадрів, на яких було виявлено об'єкт з одним і тим самим ідентифікатором, можна приблизно порахувати кількість його появи у зоні огляду камери. Для мінімальної різниці у 400 кадрів (20 секунд) найчастіше виявленою є людина, зображена на рисунку 4.10, на рисунку 4.11 зображено приклади хибних виявлень для її ідентифікатору.



Рисунок 4.10 — Найчастіше виявлена людина



Рисунок 4.11 — Хибні детекції для ідентифікатора найчастіше виявленої людини

## **Висновки до розділу 4**

Було застосовано методи аналізу відеозаписів для вихідного відеоматеріалу та отримано первинну статистику.

Алгоритми виявлення людей на основі нейронних мереж YOLOv3 та Faster R-CNN ResNet-50 FPN показують достатньо високу точність на частинах відеозапису, де люди знаходяться не щільно один біля одного, проте не виявляють об'єкти у щільно заповнених середовищах.

Відстеження об'єктів стабільно працює при пропуску об'єкта на кількох послідовних кадрах, однак при швидкому русі об'єктів та накладанню їх один на одний присвоює хибні ідентифікатори об'єктам.

Реідентифікація об'єктів показує високі результати для фронтального вигляду об'єктів, проте має складнощі з розпізнаванням бічних та задньої сторони людини та з розпізнаванням візуально схожих об'єктів.

## ВИСНОВКИ

В результаті роботи було досліджено алгоритми реідентифікації осіб, обґрунтовано вибір нейронних мереж для виявлення та відслідковування осіб, розроблено програмну реалізацію для первинного аналізу матеріалів системи відеоспостереження.

Обрані архітектури нейронних мережі для виявлення людей показують достатньо високу точність, проте для роботи вимагають використання графічних прискорювачів та не дозволяють роботу в реальному часі при виявленні об'єктів на кожному кадрі. Відслідковування і надання ідентифікаторів не є точним, оскільки на вихідних даних присутнє накладання багатьох об'єктів, що суттєво ускладнює роботу трекера та реідентифікатора.

Застосування методів повторної ідентифікації не є можливим без суттєвих обмежень до вихідного відеоматеріалу. Правильне розміщення камер, калібрування камер, параметри зйомки мають бути завчасно і ретельно продумані, проте зазвичай це не є можливим через ряд факторів. Робота з відео, отриманих з різних відеокамер, має специфічні особливості і вимагає особливий підхід до характеристик кожного відеоматеріалу.

Реідентифікація осіб суттєво полегшує обробку відео даних, проте не може гарантувати надання надійних ідентифікаторів, що не дає можливість повністю автоматизувати аналіз відео, однак надає інструменти для отримання загальної статистики.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Bedagkar-Gala A. “A survey of approaches and trends in person re-identification” [Текст] / Apurva Bedagkar-Gala, Shishir K. Shah // 2014
2. Murino V. “Re-identification: State of the Art and Current Trends” [Текст]/ Vittorio Murino // 2017
3. Wang H. “Person Re-Identification in Identity Regression Space” / Hanxiao Wang, Xiatian Zhu, Shaogang Gong, Tao Xiang // 2018
4. Hussain S. “Machine Learning Methods for Visual Object Detection” [Текст] / Sabit Ul Hussain // 2012
5. Zou Z. “Object Detection in 20 Years: A Survey” [Текст] / Zhengxia Zou, Zhenwei Shi, Yuhong Guo, Jieping Ye // 2019
6. Everingham M. “The PASCAL Visual Object Classes (VOC) Challenge” [Текст] / Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, Andrew Zisserman // 2009
7. Understanding the mAP Evaluation Metric for Object Detection [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступу : <https://medium.com/@timothycarlen/understanding-the-map-evaluation-metric-for-object-detection-a07fe6962cf3>
8. Chung A. “Cloud Computed Machine Learning Based Real-Time Litter Detection using Micro-UAV Surveillance” [Текст] / Ashley Chung, Sean Kim, Ethan Kwok, Michael Ryan, Erika Tan, Ryan Gamadia // 2018
9. Zhu J. “Online Multi-Object Tracking with Dual Matching Attention Networks” [Текст] / Ji Zhu, Hua Yang, Nian Liu, Minyoung Kim, Wenjun Zhang, Ming-Hsuan Yang // 2018
10. CS231n: Convolutional Neural Networks for Visual Recognition [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступу : <http://cs231n.github.io/neural-networks-1/>
11. Lin T. “Focal Loss for Dense Object Detection” [Текст] / Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar / 2018

12. Redmon J. “YOLOv3: An Incremental Improvement” [Текст] / Joseph Redmon Ali Farhadi / 2018
13. Reproducing training performance of YOLOv3 in PyTorch [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступу : <https://mc.ai/reproducing-training-performance-of-yolov3-in-pytorch-part-0/>
14. Bewley A. “Simple Online And Realtime Tracking” [Текст] / Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft // 2017
15. Wojke N. “Simple Online And Realtime Tracking With A Deep Association Metric” [Текст] / Nicolai Wojke, Alex Bewley, Dietrich Paulus // 2017
16. Zhang X. “AlignedReID: Surpassing Human-Level Performance in Person Re-Identification” [Текст] / Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, Jian Sun // 2018
17. Zhang Z. “A Flexible New Technique for Camera Calibration” [Текст] / Zhengyou Zhang // 1998
18. Intel Distribution of OpenVINO Toolkit [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу : <https://software.intel.com/en-us/openvino-toolkit>
19. COCO: Common Objects in Context [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступу : <http://cocodataset.org/>
20. Растрова графіка [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступу : [https://stud.com.ua/97264/informatika/rastrova\\_grafika](https://stud.com.ua/97264/informatika/rastrova_grafika)